

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Pembelajaran mesin merupakan salah satu bidang teknikal yang terus berkembang pesat saat ini, yang merupakan bagian dari kecerdasan buatan [1]. Kecerdasan buatan dapat digunakan untuk memecahkan berbagai masalah, mulai dari permasalahan bisnis, robotika, bahasa alami, matematika, analisis finansial, penalaran, dan sebagainya [2]. Salah satu pengaplikasian pembelajaran mesin adalah pada *Natural Language Processing* atau NLP, yang merupakan proses mesin untuk mempelajari sintaks mengenai bahasa manusia yang kemudian diproses dan di-*output* ke pengguna agar membentuk sebuah interaksi yang efektif dan efisien [3]. Oleh karena itu, NLP memiliki peranan penting untuk membantu membangun model, memproses informasi, dan memanipulasinya sesuai algoritma yang ditentukan dalam bentuk suara maupun teks, yang tentunya sangat bermanfaat pada berbagai bidang khususnya jurnalistik.

Pada bidang jurnalistik khususnya di era digital yang serba cepat ini, tingginya tuntutan akan akses berita terkini membuat tim jurnalistik suatu media massa perlu mengumpulkan informasi terkait suatu berita dan menyampaikannya dengan cepat agar dapat diakses oleh konsumen, terutama ketika terdapat sebuah *breaking news*. Tuntutan tersebut seringkali menyebabkan proses penulisan yang dilakukan secara terburu-buru karena harus memenuhi tenggang waktu untuk penyajian berita, sehingga waktu yang tersedia untuk memeriksa dan menyunting berita sangat terbatas dan kerap kali menyebabkan adanya kesalahan kebahasaan [4]. Kesalahan bahasa yang umum terdapat pada media massa meliputi kesalahan ejaan, kalimat (sintaksis), dan semantik (makna) [5]. Penggunaan kaidah kebahasaan yang baik dan benar pada media massa merupakan salah satu bentuk pelestarian dan pembinaan Bahasa Indonesia [6, 7], sehingga detail kecil seperti ejaan dan kaidah kebahasaan lainnya sangat perlu untuk diperhatikan agar sesuai dengan kaidah yang berlaku.

Salah satu kesalahan yang umum ditemui pada berita adalah mengenai peluluhan kata. Peluluhan yang merupakan subset dari morfofonemik, merupakan hasil dari pertemuan antar morfem (kata atau suku kata) sehingga menyebabkan perubahan fonem (huruf) agar pelafalannya menjadi lebih mudah [8]. Adapun

kaidah kebahasaan yang benar untuk peluluhan kata dasar dengan gugus konsonan (huruf konsonan ganda) adalah jika imbuhan adalah *me-* dan *pe-* maka kata dasar tersebut tidak akan luluh, sedangkan pada kata dasar dengan konsonan tunggal akan luluh [9].

Untuk mengatasi kesalahan pengejaan tersebut, maka dikembangkan sebuah aplikasi bernama U-Tapis, yang merupakan aplikasi untuk mendeteksi dan mengoreksi kesalahan kaidah kebahasaan untuk Bahasa Indonesia yang dapat membantu jurnalis [10, 11]. Pendekatan komputasional digunakan untuk membangun U-Tapis sehingga dapat menyelesaikan permasalahan yang bersifat kompleks dan memiliki informasi dari data yang besar [10]. Sebelumnya, fitur pendeteksian dan pengoreksian peluluhan kata sudah diteliti menggunakan algoritma *Jaccard Similarity*, tetapi diperoleh *F1 Score* sebesar 66.6% dari hasil evaluasinya [12]. Penelitian serupa yang menggunakan penggabungan algoritma *Damerau-Levenshtein Distance* dan BERT juga dilakukan, tetapi dalam implementasinya belum mampu memperbaiki peluluhan kata yang memiliki konsonan ganda [13]. Oleh karena itu, penelitian ini berfokus untuk mengembangkan fitur peluluhan kata agar dapat mendeteksi kesalahan penulisan peluluhan kata dengan konsonan ganda maupun konsonan tunggal menggunakan algoritma *Damerau-Levenshtein Distance*. Algoritma *Damerau-Levenshtein Distance* dipilih sebagai algoritma untuk menghitung jarak *string* target dengan *string* sumber karena merupakan pengembangan dari algoritma *Levenshtein Distance* yang memperhitungkan transposisi simbol yang berdekatan [14]. Dengan demikian, algoritma ini cocok digunakan dalam penelitian untuk membangun mesin pengoreksi ejaan, yaitu dengan membandingkan kata yang terdeteksi terhadap kamus berisi bentuk luluh yang benar dan mengembalikan nilai jaraknya sebagai dasar koreksi.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka penelitian ini berfokus untuk merancang dan membangun model pembelajaran mesin untuk mendeteksi kesalahan peluluhan kata, yang nantinya model ini akan diterapkan pada aplikasi U-Tapis. Penelitian ini diharapkan dapat membantu jurnalis dalam memeriksa dan menyunting artikel berita yang telah ditulis.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah tersebut, maka dapat dirumuskan masalah sebagai berikut:

1. Bagaimana mengimplementasikan algoritma *Damerau-Levenshtein Distance* pada *website* U-Tapis untuk mendeteksi dan mengoreksi kesalahan peluluhan kata?
2. Bagaimana mengukur akurasi, presisi, *f1-score*, dan *recall* algoritma *Damerau-Levenshtein Distance* pada *website* U-Tapis dalam mendeteksi dan mengoreksi kesalahan peluluhan kata?

1.3 Tujuan Penelitian

Adapun tujuan yang hendak dicapai dari penelitian ini berdasarkan rumusan masalah di atas yaitu:

1. Mengimplementasikan algoritma *Damerau-Levenshtein Distance* pada *website* U-Tapis untuk mendeteksi dan mengoreksi kesalahan peluluhan kata.
2. Mengukur akurasi, presisi, *f1-score*, dan *recall* algoritma *Damerau-Levenshtein Distance* pada *website* U-Tapis dalam mendeteksi dan mengoreksi kesalahan peluluhan kata.

1.4 Urgensi Penelitian

Berdasarkan latar belakang masalah yang telah dipaparkan, maka urgensi penelitian ini adalah seringnya terjadi kesalahan penulisan peluluhan kata yang dilakukan oleh jurnalis saat menulis berita karena tenggat waktu yang mendesak khususnya saat ada *breaking news*, sehingga perlu dibuat sebuah aplikasi yang dapat mendeteksi dan mengoreksi kesalahan penulisan tersebut agar penulisan dan penyuntingan berita menjadi lebih efisien dan efektif.

1.5 Luaran Penelitian

Adapun luaran penelitian ini adalah sebagai berikut.

1. API lokal untuk aplikasi berbasis *web* bernama U-TAPIS yang dapat mendeteksi dan mengoreksi kesalahan penulisan peluluhan kata.
2. Hasil evaluasi terhadap akurasi algoritma *Damerau-Levenshtein Distance* dalam mendeteksi dan mengoreksi kesalahan penulisan peluluhan kata sehingga dapat dijadikan acuan penelitian berikutnya.

3. Artikel ilmiah pada jurnal nasional terakreditasi.

1.6 Manfaat Penelitian

Adapun manfaat dari penelitian ini meliputi:

1. Menambah wawasan mengenai pengimplementasian algoritma *Damerau-Levenshtein Distance* dalam mendeteksi dan mengoreksi kesalahan penulisan kata luluh.
2. Membantu tim jurnalistik dalam memeriksa ejaan kata yang luluh secara efektif dan efisien.
3. Menambah pengetahuan mengenai performa algoritma *Damerau-Levenshtein Distance* terhadap pendeteksian dan pengoreksian kesalahan peluluhan kata, sehingga dapat menjadi rujukan untuk penelitian selanjutnya.

